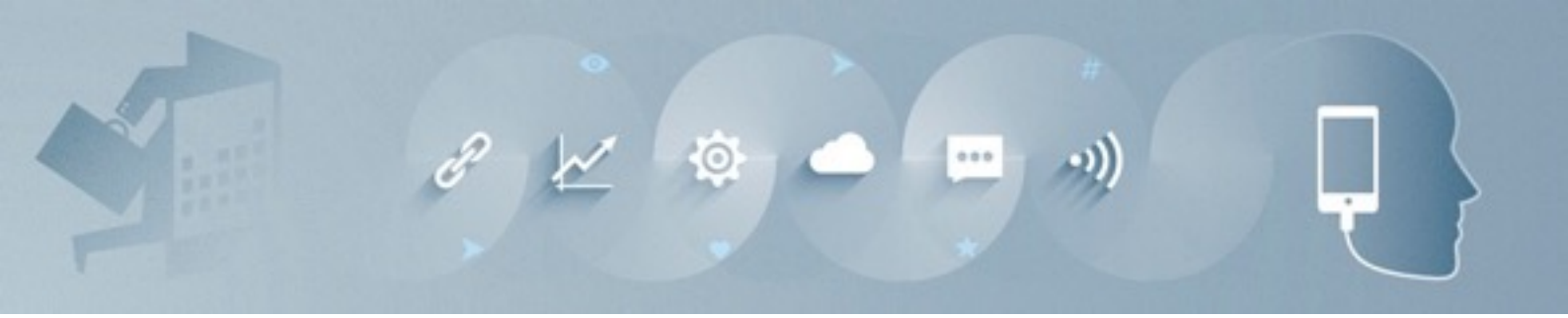

CHEMNITZER INTERNET- UND TECHNIKSOZIOLOGIE: WORKING PAPERS

Working Paper 2016-01

What is the Hidden Web?

The development, characteristics and social significance of anonymous communication on the hidden web

Christian Papsdorf

TECHNISCHE UNIVERSITÄT
CHEMNITZ

Christian Papsdorf

What is the Hidden Web? The development, characteristics and social significance of anonymous communication on the hidden web

Chemnitzer Internet- und Techniksoziologie: Working Papers 2016-01

ISSN 2367-296X

Jun.-Prof. Dr. Christian Papsdorf
Institut für Soziologie
Professur für Techniksoziologie
TU Chemnitz
Thüringer Weg 09
09126 Chemnitz
christian.papsdorf@soziologie.tu-chemnitz.de
+49371/53138163

Christian Papsdorf ist Juniorprofessor für Techniksoziologie mit dem Schwerpunkt Internet und Neue Medien an der TU Chemnitz.

christian.papsdorf@soziologie.tu-chemnitz.de

Abstract: More than two-and-a-half million people currently use the Tor network to communicate anonymously via the Internet and gain access to online media that are not accessible using standard Internet technology. This sphere of communication can be described as the hidden web. In part because this phenomenon is very recent, the subject has scarcely been studied in the social sciences. It is therefore the purpose of this paper to answer four fundamental questions: What is the hidden web? What characterises the communication sphere of the hidden web in contrast to the “normal Internet”? Which reasons can be identified to explain the development of the hidden web as a new communication sphere? And, finally, what is the social significance of the hidden web?

Zusammenfassung: Über zweieinhalb Millionen Menschen nutzen gegenwärtig das Tor Network, um anonym über das Internet zu kommunizieren und Zugriff auf Online-Medien zu erhalten, die mit gewöhnlicher Internettechnik nicht nutzbar ist. Diese Kommunikationssphäre kann als Hidden Web bezeichnet werden. Unter anderem weil es sich um ein sehr junges Phänomen handelt, liegen bisher nahezu keine sozialwissenschaftlichen Erkenntnisse zu dem Thema vor. Dementsprechend werden hier vier grundlegende Fragen beantwortet: Was ist das Hidden Web? Welche Eigenschaften weist die Kommunikationssphäre des Hidden Web im Vergleich zum „normalen“ Internet auf? Welche Gründen lassen sich identifizieren, die die Entstehung des Hidden Web als neue Kommunikationssphäre erklären können? Und welche gesellschaftliche Bedeutung kommt dem Hidden Web schließlich zu?

Inhalt — Contents

1	Introduction	5
2	Linguistic differentiation of the hidden web and an overview of the literature	6
3	Characteristics of communication via the hidden web	8
4	The creation of the hidden web as a response to the development of the visible web	11
5	The social significance of the hidden web	15
6	Summary and prospects	17
	Literature	18

1 Introduction

The World Wide Web and many Internet media based upon it have enjoyed great popularity for many years. Shopping platforms, blogs, wikis, news sites and social networks are used throughout the world and it is no longer possible to imagine everyday life without them (Blank & Groselj, 2014). And yet there have always been the hidden services of the so-called hidden web (Bergman, 2000) alongside the publicly accessible and freely addressable media offerings. For a long time they remained limited to private contacts, members of particular organisations or other equally tightly circumscribed groups of people.

In the more recent past the hidden web has continued to develop, however, so that the ways in which it is used have come to resemble those of the commonly known side of the Internet. On the hidden web, too, there are now social networks, wikis and, of course, websites. What differentiates this content, however, is that it is accessible only through specific software and that it cannot, for example, be captured by search engines. The programs necessary to access the hidden web, such as the popular Tor browser (Li et al., 2013, p. 1269), guarantee a considerable degree of anonymity using encryption technology. It is therefore impossible, or extremely difficult, to identify users. Engaging in this form of communication provides protection from the state surveillance of "classic" Internet communication, which has been omnipresent for several years, and from the aggregation of user data by Internet companies. In combination with other services, such as the decentralised online currency Bitcoin, the hidden web is developing into a much-used alternative to the "standard" Internet. Approximately two-and-a-half million people now use the Tor network (Tor Metrics, 2015a), with the greatest number of users resident in the USA (17%), Germany (9%), Russia (8%), France (6%) and Great Britain (4%) (Tor Metrics, 2015b). A large proportion of users also lives in countries in which access to the Internet is restricted (Fowler, 2012). There are, furthermore, more than 30,000 distinct .onion addresses, each of which belongs to a website or other service on the hidden web (Tor Metrics, 2015c). A marketplace (for both legal and illegal products) that was at one time popular but has since then been shut down was Silk Road. Even in the early days it boasted a turnover of US$ 15 million, which must be interpreted as an indication of extensive use (Christin, 2013).

An online communication sphere has thus been created that is characterised by public communication, as in the form of mass media, but that simultaneously isolates itself from the publicness of the "classic Internet" by being accessible only through the use of specific software. This new form of Internet communication has been the subject of neither theoretical reflection nor empirical research in the social sciences. The present article therefore aims to develop a first assessment of this subject. The next section starts by exploring the terms with which this communication sphere can be discussed and what is already known about it. The third section

examines the characteristics of communication via the hidden web as they contrast with the standard Internet. The fourth section develops the theory that the hidden web can be seen as a reaction to the increasing standardisation and commercialisation of the Internet as a whole. The fifth section focuses on the social significance of the hidden web.

2 Linguistic differentiation of the hidden web and an overview of the literature

There is at present no consistent definition of the terminology associated with the use of the hidden web. The communication services relevant to this discussion are referred to as the hidden web, deep web, invisible and darknet. The corresponding terms clearnet, visible web, indexable web and surface web are used to describe classic Internet media and communication. These terms are not scientifically founded and are often used in a normative sense, for example by referring to all websites used by terrorists as the dark web (Chen et al., 2008). For the discussion that follows, it is proposed that "the" Internet should be divided into two terms, the hidden web and the clearnet as these best reflect the central complex of criteria, namely publicness and visibility.

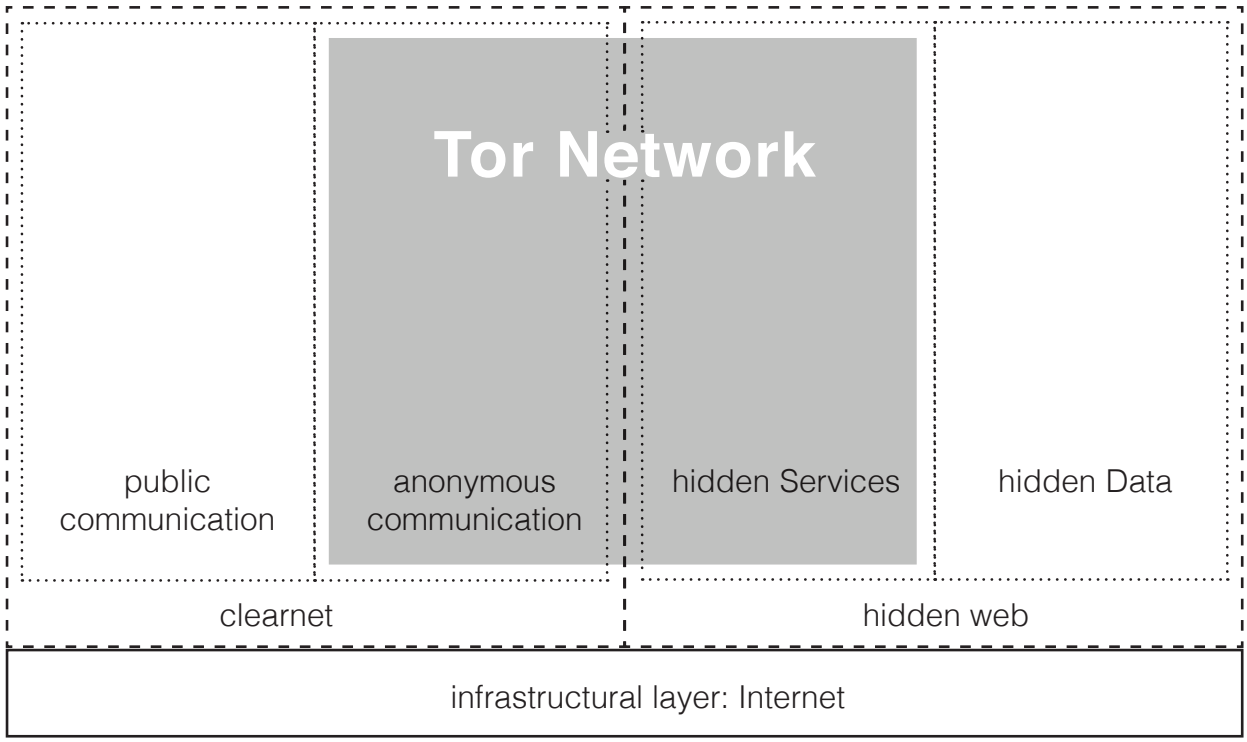

Figure 1: Overview of the hidden web and clearnet. Diagram: Christian Papsdorf

The use of these two separate terms assumes that both communication spheres share the same basic infrastructure: the Internet. The Internet, with its physical data pipelines, servers and routers and the countless software protocols, forms the foundation of a plethora of web media, which can subsequently be divided into two communication spheres. The hidden web can then be further

subdivided into two forms of media. To this end, McCoy et al. (2008) differentiate between non-interactive and interactive forms of use. This differentiation is essential with reference to the transformation of the hidden web mentioned above. For a long time, non-interactive forms of use, such as databases, file-hosting services and peer-to-peer file sharing constituted the dominant forms of use. Since the Tor network was launched ten years ago, however, the number of online media that permit direct (and in some cases real-time) communication between users has increased. This differentiation will be reflected in the terms “hidden data” and “hidden services”. The communication made possible by hidden services will be at the heart of this discussion. These hidden services can be accessed only by users of the anonymous Tor network. All media available on the clearnet can also be used via the Tor network, however, so that the Tor network cannot be equated with the hidden web despite the fact that the former is a necessary condition for the latter. Figure 1 visualises this differentiation of Internet communication.

The current state of research on the hidden web in the social sciences can only be described as deficient. Until now, phenomena beyond the clearnet have been the subject of very little theoretical examination or empirical research. In computer science and the engineering sciences, too, there are very few studies of the hidden web. And so it continues to be true that “surprisingly, very few researches shed light on such an anonymizing network“ (Chaabane et al., 2010, p. 167).

Research into the hidden web in the field of computer science concentrates primarily on the development of web crawlers so that they can capture the sites located on the deep web (Zheng et al., 2013, p. 801; Raghavan/ Garcia-Molina, 2000; Ntoulas et al., 2005; Barbosa/ Freire, 2007, p. 441; Hedley et al., 2006, p. 213, Ipeirotis et al., 2001, p. 67). Bergman (2000) uses a crawler such as this to show that the hidden web was, at the time that the research was carried out, approximately 500 times bigger than the freely accessible Internet, containing more than 200,000 websites and growing at a faster pace than the clearnet. A large proportion of the information available on the hidden web grew from databases on particular subjects, internal organisational networks and journalistic publications: hidden data, in other words, rather than hidden services. There are also individual studies about the degree of anonymity (González, 2013, p. 73), about the data transfer rate (Liška et al., 2010, p. 542), about the localisation of “hidden servers” (Øverlier/ Syverson, 2006, p. 100) and about possible protections against attacks (Castillo-Pérez/ Garcia-Alfaro, 2013, p. 197). These studies, which have a strong focus on technological aspects, are supplemented by a small number of results from other research, some of it in the field of the social sciences.

Chen et al. (2008, p. 1347) regard the hidden web (or, to use their terminology, the dark web) as the other side of the Internet, which is used by terrorists and extremists for their activities. They show that there are websites through which the various actors spread their ideologies and create

networks. Nazemi (2012, p. 855), on the other hand, argues with reference to legal aspects that the Tor network can make an important contribution to the process of democratisation in countries with limited freedom of the press. On the other hand, he argues, the Tor network's substantial anonymity prevents the legal prosecution of copyright infringements resulting from file sharing. Forte (2006, p. 85) emphasises the ways in which the Tor network can be used by criminals. In an overview study Li et al. (2013, p. 1269) show that the Tor network is used to varying degrees in different countries. The largest number of servers and volunteers who work on the project are located in Germany and the USA. They also found that the Tor network is by far the most frequently used of all the services that provide anonymity. With the exception of unique forms of use, little is known about the hidden web in the social sciences. The following section therefore consists of a discussion of its fundamental characteristics.

3 Characteristics of communication via the hidden web

Although very little is known at present about the hidden web, it is possible to discuss the fundamental characteristics of communication via hidden services. Self-representations by the various organisations (primarily by the Tor network), reports published by the media and the first explorative online ethnographic study reveal that there are six basic differences between the hidden and the clearnet. These refer to the actors involved in, the media available on and communication via the hidden web.

Criterion	Hidden web	Clearnet
Users	2.5 million; homogenous group	More than 3 billion; heterogeneous group
Media	Extent and degree of differentiation: medium; small degree of hierarchical organisation	Very extensive, highly differentiated; hierarchical
Anonymity and data protection	Technologically guaranteed for users and services	Low
Publicness and obscurity	Technologically and socially enabled	Only socially enabled
Legal regulations	Difficult to enforce	Almost entirely enforceable
Usability	Low	High

Table 1: Comparison of the characteristics of the clearnet and the hidden web

With respect to users, two characteristics are of particular significance: the number and their composition. At present a total of 40 percent of the world's population, or more than three billion people, uses the Internet (Internet Live Stats, 2015). More than two-and-a-half million people now use the Tor network (Tor Metrics, 2015a). The group of users who can use hidden services is therefore not insignificant or negligible. And yet, at approximately 0.1 percent of the total number of users, it is a very small group when compared to the users of the clearnet. Use of the clearnet and the composition of its users varies considerably between countries (Çilan et al., 2008). In Western societies, users of the Internet have become a highly heterogeneous group characterised by a very broad spectrum in terms of metrics such as age, degree of education, income, political and social values (Zillien, 2009, p. 153).

Use of the Tor network, by contrast, must fulfil two important preconditions. Firstly, users much have a specific interest in using the software and hidden services. As the most important innovation of the Tor network lies in the high degree of anonymity guaranteed by technological means it is used primarily by people who value data protection in the face of the widespread surveillance of Internet communication. Secondly, users must be willing and able to use the unconventional technology. The installation of the software and discovery of hidden services requires a higher degree of dedication (and, in some cases, specialist knowledge) than use of the clearnet. Current users are therefore considered to be early adopters (Rogers, 1962) who are interested in technology and sensitive to data protection.

The media available on the clearnet are now many and varied. They consist of technologies rooted in the Internet (World Wide Web, e-mail and data transfer), digitised classic media (IP telephone calls and TV) and a large number of combinations. The variety of digital media has become incalculable as a result of the great innovativeness and rapidly advancing technology. New platforms, social networks and digital mass media are constantly being created, particularly in the form of apps for user devices. The media landscape of the clearnet is therefore extensive, differentiated and also hierarchical because individual providers of particular media categories (streaming, cloud storage, e-mail, video platforms, social networks, microblogs, sales platforms, search engines, online encyclopaedias) have in many cases achieved dominant market positions.

According to estimates, there are more than 30,000 hidden services online, covering a variety of media (Tor Metrics, 2015c). These include wikis, chat and e-mail programs, search engines, websites, sales platforms, blogs, social networks (Biryukov et al., 2014) and whistle-blowing services. The hidden web is therefore characterised both in quantitative and structural terms by a widespread range of media. Only in terms of user numbers and the newness of this communication sphere does it lag far behind the clearnet. Online black markets are the most well-known service of the hidden web, not least because they have provoked considerable public interest. They exemplify

the fact that no provider has been able to acquire a dominant position comparable to that of eBay or, later, Amazon, as legal marketplaces on the clearnet (Power, 2014).

The third set of differences results from the technological particularities of the hidden web and refers primarily to qualitative aspects of communication. The main difference lies in the fact that online communication via the Tor network guarantees a high degree of anonymity, whereas this is generally not the case on the clearnet. Internet anonymity in this respect refers to the impossibility of connecting communications in which one person engages in such a way that their identity is revealed (Federrath/ Pfitzmann, 1998, p. 319). In addition, the use of pseudonyms has become controversial on the clearnet in recent years. Both media organisations and political figures demand that real names be used (Boyd, 2012). Beyond the use of pseudonyms, users of the clearnet are easily identifiable via their IP addresses as well as various tracking options (such as cookies). The advantages and disadvantages of anonymous Internet use will not be judged here, although it will be stated that users have an increasingly strong interest in data protection (Initiative D21, 2013, p. 13). On the hidden web anonymity is enabled by technology by transporting encrypted and re-encrypted data packages via constantly changing routes and detours. Inasmuch as this process functions as intended it is virtually impossible to identify users. The providers of hidden services, too, are difficult to identify. They can be hosted on dedicated servers in a decentralised manner and their URLs, unlike those on the clearnet, do not have to be registered and are not coordinated by an organisation (such as ICANN).

Another important difference lies in the degree of publicness. Even with respect to the clearnet it is no longer always easy to clearly delineate private and public communication, as one can with respect to offline media. This is because users of many media (such as social networks) can themselves manage the degree of publicness of their communication (Papsdorf, 2013, p. 211). The question of the relationship between the public and the private cannot be definitively answered for online media but instead depends on their visibility. Selinger and Hartzog (2014) introduce the term “obscurity” to this end. In the online world, publicness exists wherever obscurity, which is fed by numerous sources, gives way to visibility. In the clearnet obscurity exists when media of individual communication are used, when the group of recipients is consciously limited within communities (such as social networks) and when communication platforms cannot generate sufficient awareness because they are difficult to find or difficult to understand. These sources of non-publicness can also be identified with respect to the hidden web. In the case of the hidden web they are joined by a further source of non-visibility, however, which is also in the final analysis reflected in the term “hidden web”. This state of being concealed encompasses three levels. Firstly, the Tor network is itself not yet known to the wider public but is instead hidden. Secondly, the hidden services are not easy to find because there are no indexes

and search engines capture only a fraction of them. Thirdly, the providers and users are visible only to a limited extent in terms of the anonymity described above. In contrast to the clearnet, the users of the hidden web are not in sole control of privacy and publicness. Instead, visibility is already triply limited by technological means.

Another difference relates to the legal aspects of Internet use. Although some may assume that the Internet is an extralegal space, online communication is decidedly and extensively subject to legal norms. Despite questions of overlapping but divergent national regulations, there are no longer any legal grey zones. The ability to enforce existing laws (Papsdorf, 2013, p. 159) alone is in some cases problematic. As a result of the technologically enabled anonymity the assignment of infringements to individual culprits is sometimes a very complicated process on the hidden web. The consequences of this state of affairs are multifaceted. On the one hand, the Tor network can make an important contribution to the process of democratisation in countries in which the freedom of the press is limited because it makes both the persecution of members of the opposition and the censorship of opinions much more difficult. This is one reason for which the social network Facebook can also be accessed via the hidden web. On the other hand, the Tor network prevents the criminal prosecution of copyright infringements, fraud and more serious crimes and offences (Nazemi, 2012, p. 855).

The media available on the hidden web also differ from those on the clearnet in terms of their usability. User-friendliness has steadily increased on the clearnet as a result of hardware improvements, more powerful software and, most importantly, strong competition between various online media. On the hidden web, by contrast, user-friendliness is low. Users of the hidden web must be able to overcome the various barriers to entry into and use of the aforementioned services. They must themselves research cryptic .onion URLs and work with low data transfer rates (because of encryption). In part because of the small number of users and concomitantly lower potential profits, the media available on the hidden web are, furthermore, less technologically elaborate than their counterparts on the clearnet, which in some cases are associated with generous budgets.

4 The creation of the hidden web as a response to the development of the clearnet

After juxtaposing the characteristics of communication on the hidden web with those of the clearnet in the preceding section, the discussion that follows will consider this from another angle. It is assumed that the reasons for which the hidden web developed can be understood only in relation to the development of the clearnet. The theory that the hidden web should be regarded as a response to the development of the clearnet will be discussed below. It is believed that the clearnet became less attractive to some user groups from the late 1990s onwards as a result of its

popularisation, commercialisation, institutionalisation and because of increasing surveillance. Demand for a new communication sphere that does not share these disadvantages was born and, eventually, satisfied. With reference to the characteristics of the hidden web discussed in the preceding section it will be shown that in some respects it represents a newer variation of the early clearnet of the 1990s and, in this sense, though with different means, recreates the communication platforms of the past. In what follows the six dimensions will be discussed with respect to their transformation within the framework of communication on the clearnet and ways in which the hidden web compensates for these changes.

The first complex of criteria encompasses the number and composition of users. In the early stages of the clearnet users were a relatively small elite of technological visionaries and developers as well as decision-makers (Kirsten, 1999; Campbell-Kelly/ Garcia-Swartz, 2013). At this point there was little difference between users and developers. Early users generally developed “their” Internet in a cooperative and collaborative manner without interference from external influences. By 1995 as many as 15 million people were online (Internet World Stats, 2015), however. This resulted in a plethora of professional media organisations and companies that took charge of the continued development of the web and used various means to make the new users into users of their media. The early developers and activists were not equipped to fend off such powerful competition and thus lost influence and retreated into their niches (such as the open-source arena).

The current hidden web, in contrast, like the clearnet of 20 years earlier, is a communication sphere that can be shaped at will by technological pioneers. As a result of the small number of other users, the early adopters (Rogers, 1962) are not in competition with media organisations and, in contrast to the present-day clearnet, do not need venture capital in order to develop or popularise a communication platform (Chang, 2004). Instead, there is a sense of new beginnings and a sense of “anything goes”.

The aspect of malleability is also central when it comes to the various media. When it comes to the clearnet it has become apparent that a stable system of media (with dominant market leaders in many spheres) has established itself despite the fact that rapid innovation continues apace. On the hidden web, on the other hand, this is not the case in virtually any field. In contrast to the clearnet of the early 1990s, media providers do not have to be invented, however. They must simply be transferred to the hidden web under anonymous conditions. Whenever the transfer does not function seamlessly the necessity to create media innovations arises. And so decentralised, virtual payment systems (such as Bitcoins) are employed to transfer money anonymously and search engines that do not save any of their users’ data are used. On the

hidden web these media innovations enable and feed one another, as in an ecosystem. The combination of various services is essential to the meaningful use of anonymous communication.

The next element, anonymity and the data protection that it implies, was also a de facto aspect of the early clearnet. On the one hand, early developers had a (surprisingly) keen interest in privacy (Braman, 2011, p. 800) and, on the other hand, users were largely immune to surveillance as a result of the novelty and the initially widely underestimated significance of information and digital communication. Technologies that enable the extensive storage and analysis of communication were not yet widespread. Furthermore, neither state nor privately held organisations had any idea what to do with online communication. This changed radically, at the latest with the launch of PRISM in 2007, and of Tempora shortly thereafter. It also became possible to show that the analysis of user data constituted the foundation of the business models of particular online media (such as social networks) as well as hardware producers (Fuchs, 2012). Communication on the hidden web, on the other hand, is comprehensively protected from surveillance by its technological architecture, as discussed above. This signifies the targeted and active creation of data protection, in contrast to the early clearnet. The terms in which the Tor network describes itself show very clearly that differentiation from the clearnet has been a central motif of its development: “Tor is free software and an open network that helps you defend against traffic analysis, a form of network surveillance that threatens personal freedom and privacy, confidential business activities and relationships, and state security” (Tor Project, 2015).

The relationship between publicness and privacy in the early phase of the clearnet was ambiguous. Although many mass media, such as Usenet (which was popular for a long time), were used, this communication barely reached the wider population beyond what was still a small Internet. In comparison to later developments, communication on the early clearnet should be understood as largely concealed, and public only to a very limited degree. The clearnet has become a permanent fixture of the media landscape (Hasebrink et al., 2013). Internet communication also has an increasing influence on classic media as communication originating on the Internet (as, for example, on Twitter) is disseminated via the television, radio and written publications (Papsdorf, 2013, p. 225). Apparently private communication, too, repeatedly finds its way into the public sphere, so that some have already declared the dawn of the age of post-privacy (Heller, 2011). This has resulted in a pressure to conform (Rosander/ Eriksson, 2012), which is evidenced by nettiquettes, rules of conduct (in the general terms and conditions of various media) and the professionalization of users (as exemplified by e-mail signatures). As discussed in the third section, communication via the hidden web is characterised by a triple obscurity. This makes it possible to avoid visibility despite its fundamental publicness.

Another parallel to be discussed in this context is related to the enforceability of legal regulations. The myth of the Internet as an extrajudicial space grew out of the early days of the clearnet, in which some areas of online communication were not subject to legal norms and law enforcement agencies were not in a position to prosecute crimes and offences effectively. The freedom that accompanied this state of affairs was then limited in favour of a higher degree of traceability of crimes and offences. The example of data retention shows that even online communication that bears no connection to illegal activities is now subject to legal surveillance. The hidden web has made it possible to recreate this freedom in terms of the inability to censor communication and the inability to sanction legal infringements. Whereas this was a coincidental side effect of the early clearnet it is a consciously developed characteristic of the hidden web.

The parallels between the early clearnet and the hidden web discussed above can be regarded as a reaction to the development of the clearnet. The issue of usability shows that there are also unintended (and unfavourable) parallels. In the early 1990s the Internet could be used only by people who had in-depth knowledge of hardware and software, not least so that they could resolve the constantly occurring connection problems, programing faults and other issues. Over the course of the years the technology became increasingly easy to use. By the turn of the twentieth to the twenty-first century users needed only an average degree of knowledge in order to communicate online. Nowadays the clearnet can be used with almost no prior knowledge, without complicated configurations, quickly and efficiently, and it is also accessible to older people, for example. Apps in particular, which are available on mobile user devices, can generally be used immediately and with no difficulties at all (Barnard, 2013). In addition, many media have become increasingly user friendly throughout the course of their existence, in part in order to survive the competition with other providers. As discussed in the third section, the hidden web is characterised by relatively low usability. The causes for this could, as with the clearnet, be remedied with increasing use, however. This is, therefore, primarily a transitional effect.

The theory that the hidden web can be viewed as a response to the clearnet can to a great extent be affirmed with reference to the six differences discussed in the previous section. The hidden web appears to fill a gap that developed as the clearnet became a mainstream medium. The parallels arise from two different causes, however. On the one hand, the anonymity, the privacy of communication and the protection from censorship have been created intentionally by the leading figures in the field to differentiate the hidden web from the advanced clearnet. On the other hand, the number and composition of users, the structure of the media and usability are mainly the result of the novelty of this communication sphere and, in part, unintended. One might therefore ask whether the hidden web will undergo a development analogous to that of the clearnet. It is to be assumed that a development similar to that of the clearnet is more likely with respect to the

latter group of characteristics than it is with respect to the former group of characteristics, which could be said to constitute the core of the hidden web. Changes to the first three characteristics would render the hidden web obsolete.

5 The social significance of the hidden web

There can be no doubt that the overall social significance of Internet communication is great. Structures relating to communication and interaction have changed profoundly as a result of the Internet in many social spheres, such as education, the economy, politics, love, health and journalism. And yet a nuanced approach is necessary when assessing the significance of the Internet to society. There exist a wide variety of media, services and communication platforms, all of which have a distinct significance and different effects on the segments of society to which they relate. In some cases the consequences of Internet use are positive and at other times they are negative; they can have a democratising effect or not; sometimes they lead to concentration and sometimes they result in decentralisation. It is therefore also necessary to examine the significance to society of the communication media of the hidden web in a nuanced manner.

With respect to legal aspects, Nazemi (2012, p. 855) argues that the Tor network can make an important contribution to the process of democratisation in countries characterised by a limited freedom of the press. He also argues that its fundamental anonymity prevents the legal prosecution of copyright infringements resulting from file sharing. Communicating through the hidden web therefore means, first and foremost, that one cannot be prosecuted for voicing opinions or engaging in transactions or breaches of the law. This can be beneficial to society (as, for example, when it comes to revelations or the avoidance of censorship) or harmful (as when users become the victims of fraud). The recent past has shown that breaches of the law, such as those that are widespread on digital black markets on the hidden web, can now be more effectively sanctioned, also through classic “offline police work” (Ball et al., 2013). They are therefore slowly becoming less widespread. Data protection is of central importance not only in repressive regimes but for all users of the Internet, and is thus of increasing importance to 80 percent of the population of Western societies.

In Germany, concerns relating to data protection and security are the most common reasons for not using the Internet (Initiative D21, 2013, p. 13). The differentiation of different types of users (Initiative D21 2013: 44ff.) also shows that many users are establishing conservative practices. The “outside sceptic”, “careful pragmatist” and “thoughtful professional” types use only carefully selected web services and consciously reject others. Even among adolescent users ten percent can now be described as sceptics. Külcü und Henkoğlu (2014, 768) continue to show that a large proportion of

Facebook users changes the standard settings to protect personal data. There is therefore an increasing awareness that one's data needs to be protected from abuse or unwanted use.

Use of this kind can be realised in two different ways: It is possible to stop using specific services or the Internet as a whole, or one can use the services of the hidden web through the Tor network so that one's personal data cannot be accessed by third parties. In some cases, services located on the clearnet, such as the social network Facebook, are also available as hidden services for the Tor network. Whenever people do not want to eschew particular forms of Internet communication but do value communication that keeps data secure, the social significance of the hidden web becomes clear. It provides protection from surveillance and censorship of online communication through state authorities, from economic exploitation by advertisers, data merchants and profilers, and from the exposure of communication about sensitive subjects (such as illness or controversial political views) to interested organisations (such as health-insurance providers). It also protects personal data beyond the contents of online communication, including one's address, professional or marital status. Through the accumulation of data it has become easy for third parties to generate such information with recourse to only very little input. The hidden web also makes it possible for users to escape the so-called filter bubble (Pariser, 2011), which is the result of algorithms that use past online behaviour and known information to provide similar information again and again.

These possible uses imply three important social functions of the hidden web, all of which are closely connected to one another. Firstly, it enables users to protect their identities and data. Secondly, it recreates spaces in which unfettered communication is possible. Thirdly, it nurtures divergent behaviours.

As described above, the right to protect one's identity and data with respect to Internet communication in particular has in the recent past come under increasing pressure. A plethora of laws governs the right to self-determination when it comes to information about oneself, and thus also the divulgement and use of personal data. And yet a considerable quantity of online communication is subject to the surveillance and espionage of many organisations (Fuchs, 2015). The anonymity of the Tor network guarantees a large degree of protection for hidden services and their users. Personal data, both online and offline, cannot be mined for economic profit or subject to political surveillance. The hidden web thus constitutes a countermeasure that makes Internet communication attractive again, even for users concerned with data protection.

The fourth section showed that that communication on the clearnet has become increasingly standardised and thus also increasingly regulated. According to the mode of operation of a panopticon (Hope, 2005) its users' behaviour is altered by the very possibility that information about their behaviour might be saved and stored for future use in order to avoid later sanctions

(Berger et al., 2014). If this possibility is ruled out, as it is on the hidden web, users can communicate freely and without (self-imposed) constraints. Although this is also, theoretically, possible offline, offline communication is subject to spatial and temporal constraints. This is particularly pertinent with respect to sensitive or niche subjects that apply only to a small number of people worldwide or that are effective only on a transnational level.

Deviance, which can encompass behaviour that transgresses social expectations as well as legal regulations, in many cases leads to sanctions and stigmatisation. This is one of the reasons for which those who transgress norms and delinquents have an interest in anonymity, which is well served by the hidden web. Transgressive behaviour in the context of digital communication is to be found in a wide variety of contexts, including politics, sexuality, religion, the trade in goods, services, journalism, education and health. As Christin (2012) shows using the example of the no-longer-extant hidden-web marketplace Silk Road, there is demand for a broad spectrum of illegal services and, above all, goods. Here it seems to be the case that transgressive behaviour moves from the offline world onto the Internet, whereas deviant behaviours originating online, such as Internet bullying and cyber stalking (Kraft/Wang, 2010; Sabella et al., 2013), is more prominent on the clearnet.

6 Summary and prospects

Until now, very little has been known about the hidden web despite its comparatively high number of users. The present article therefore attempts to answer four fundamental questions. Firstly, it has been possible to show that one can create a linguistic distinction between the hidden web and the clearnet although both are based on the communication infrastructure of the Internet. It is also necessary to differentiate between hidden data and hidden services, whereby the latter enable communication in the narrower sense (and not only the exchange of data). It must also be noted that the Tor network is not synonymous with the hidden web because it can also be used for anonymous communication on the clearnet. Secondly, the juxtaposition of the clearnet and the hidden web shows that the latter is characterised by six attributes. The group of users is comparatively small and homogenous. The extent and degree of differentiation of the available media is more limited than that on the clearnet, and the structure of the media is less hierarchical. One central characteristic is that data protection for users and services can be guaranteed by technical means. As a result the degree of obscurity (in the sense of the relationship between the publicness and privacy of communication) is dependent not only on social but also on technical factors. Legal regulations (in the form of copyright or censorship) are difficult to enforce because of this anonymity, although enforcement is becoming increasingly efficient. In contrast to the clearnet, the hidden web is characterised by low usability in several respects.

Thirdly, it was shown that the genesis of the hidden web can be interpreted as a response to the development of the clearnet. The clearnet became less attractive to certain groups of users from the late 1990s onwards as a result of its popularisation, commercialisation, institutionalisation and because of its surveillance. Fourthly, the social significance of the hidden web was examined. Inasmuch as this can be assessed at this early stage, the hidden web enables personal and data protection and the reclamation of spaces for free communication as well as facilitating transgressive behaviour.

Empirical research into the hidden web is essential, building on this primarily theoretical foundation. The perspective presented here, comparing the hidden to the clearnet, can be helpful in developing the specifics of hidden-web use. Motivations for using the hidden web, concrete modes of usage and the influence of technically guaranteed anonymity on the achievement of successful communication would be of primary interest.

Literature

Ball, J., Arthur, C./ Gabbatt, A. (2013). FBI claims largest Bitcoin seizure after arrest of alleged Silk Road founder. Retrieved from http://www.theguardian.com/technology/2013/oct/02/alleged-silk-road-website-founder-arrested-bitcoin.

Barbosa, L./ Freire, J. (2007). An Adaptive Crawler for Locating Hidden-Web Entry Points. In: WWW '07 Proceedings of the 16th international conference on World Wide Web, 441–450.

Barnard, Y., Bradley, M. D., Hodgson, F./ Lloyd, A. D. (2013). Learning to use new technologies by older adults: Perceived difficulties, experimentation behaviour and usability. In: Computers in Human Behavior 29, 1715–1724.

Berger, P. A./ Brumme, R., Cap, C. H./ Otto, D. (2013). Überwachung des digitalen Raumes. In: Soziale Welt 65, 221–246.

Bergman, M. K. (2000). White Paper: The Deep Web: Surfacing Hidden Value. In: Journal of Electronic Publishing 7. Retrieved from http://resources.mpi-inf.mpg.de/d5/teaching/ws01_02/ proseminarliteratur/ deepwebwhitepaper.pdf.

Biryukov, A., Pustogarow, I./ Weinmann, R.-P. (2014). Content and popularity analysis of Tor hidden services. In: ICDCS Workshops 2014: 188-193.

Blank, G./ Groselj, D. (2014). Dimensions of Internet use: amount, variety, and types. In: Information, Communication & Society 17, 417–435.

Boyd, D. (2012). The Politics of “real names”. Power, context, and control in networked publics. In: Communications of the ACM 55, 29–31.

Braman, S. (2011). Privacy by design: Networked computing, 1969–1979. In: new media & society 14, 798–814.

Campbell-Kelly, M./ Garcia-Swartz, D. D. (2013). The history of the Internet: the missing narratives. In: Journal of Information Technology 28, 18–33.

Chaabane, A., Manils, P./ Kaafar, M. A. (2010). Digging into Anonymous Traffic: a deep analysis of the Tor anonymizing network. In: Network and System Security (NSS) - 4th International Conference on Network and System Security, 167–174.

Castillo-Perez, S./ García-Alfaro, J. (2013). Onion routing circuit construction via latency graphs. In: Computers & Security 37, 197–214.

Chang, S. J. (2004). Venture capital financing, strategic alliances, and the initial public offerings of Internet startups. Journal of Business Venturing 19, 721–741.

Chen, H., Chung, W., Qin, J., Reid, E., Sageman, M./ Weinmann, Gabriel (2008). Uncovering the Dark Web: A Case Study of Jihad on the Web. In: ASIS&T 59, 1347–1359.

Christin, N. (2013). Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace. Retrieved from https://www.andrew.cmu.edu/user/nicolasc/publications/Christin-WWW13.pdf

Çilan, C. A., B. A. Bolat/ E. Coşkun (2008). Analyzing digital divide within and between member and candidate countries o f European Union. In: Government Information Quarterly 26, 98–105.

Federrath, H./ Pfitzmann, A. (1998). Anonymität, Authentizität und Identifizierung im Internet. In: Bartsch, M./ Lutterbeck, B. (eds.): Neues Recht für neue Medien, Informationstechnik und Recht 7 (pp. 319–328). Karlsruhe: Schriftenreihe der DGRI.

Forte, D. (2006). Advances in Onion Routing: Description and backtracing/ investigation problems. In: Digital Investigation 3, 85–88.

Fowler, G. A. (2012). Tor: An Anonymous, And Controversial, Way to Web-Surf. Retrieved from http://www.wsj.com/news/articles/SB10001424127887324677204578185382377144280

Fuchs, C. (2012). The Political Economy of Privacy on Facebook. In: Television & New Media 13, 139–159.

Fuchs, C. (2015). Societal and ideological impacts of Deep Packet Inspection Internet surveillance. Information, Communication & Society 16, 1328–1359.

González, P. C. (2013). Fingerprinting Tor. In: Information Management & Computer Security 21, 73–90.

Hasebrink, U., Schulz, W., Deterding, S., Schmidt, J.-H., Schröder, H.-D., & Sprenger, R. (2013). Leitmedium Internet? Mögliche Auswirkungen des Aufstiegs des Internets zum Leitmedium für das deutsche Mediensystem. Retrieved from http://www.hans-bredow- institut.de/webfm_send/734

Hedley, Y.-L., Younas, M., James, A. & Sanderson, M. (2006). Sampling information extraction and summarisation of hidden web databases. In: Data & Knowledge Engineering 59, 213–230.

Heller, C. (2011). Post Privacy: Prima leben ohne Privatsphäre. München: C.H. Beck.

Hope, A. (2005). Panopticism, play and the resistance of surveillance: case studies of the observation of student Internet use in UK schools. In: British Journal of Sociology of Education 26, 359–373.

Initiative D21 (2013). D21 - Digital - Index: Auf dem Weg in ein digitales Deutschland?! Retrieved from http://www.initiatived21.de/wp-content/uploads/2013/04/digitalindex.pdf. (Abruf 23.04.2015).

Internet Live Stats (2015). Internet Users. Retrieved from http://www.Internetlivestats.com/Internet-users/

Internet World Stats (2015). Internet Growth Statistics. Retrieved from http://www.Internetworldstats.com/emarketing.htm

Ipeirotis, P. G., Gravano, L./ Sahami, M. (2001). Probe, count, and classify: categorizing hidden web databases. In: SIGMOD '01: Proceedings of the 2001 ACM SIGMOD international conference on Management of data, 67–78.

Kirsten, P. T. (1999). Early Experiences With the Arpanet and Internet in the United Kingdom. In: IEEE Annals of the History of Computing 21, 38–44.

Kraft, E. M./ Wang, J. (2010). An exploratory study of the cyberbullying and cyberstalking experiences and factors related to victimization of students at a public liberal arts college. In: International Journal of Technoethics 1, 74–91.

Külcü, Ö./ T. Henkoğlu (2014). Privacy in social networks: An analysis of Facebook. In: International Journal of Information Management 34, 761–769.

Li, B., Erdin, E., Gunes, M. H., Bebis, G./ Shipley, T. (2013). An overview of anonymity technology usage. In: Computer Communications 36, 1269–1283.

Liška, T., Sochor, T./ Sochorová, H (2010). Comparison between normal and TOR-anonymized web client traffic. In: Procedia CS 01, 542–546.

McCoy, D., Bauer, K., Grunwald, D., Kohno, T./ Sicker, D. (2008). Shining Light in Dark Places: Understanding the Tor Network. In: Borisov, N./ Goldberg, I. (eds.), Privacy Enhancing Technologies (pp. 63–76). Berlin: Springer.

Nazemi, N. (2012). DMCA § 512 Safe Harbor for Anonymity Networks Amid a Cyber-Democratic Storm: Lessons from the 2009 Iranian Uprising. In: Northwestern University Law Review 106, 855–892.

Ntoulas, A., Zerfos, P./ Cho, J. (2005). Downloading Hidden Web Content. In: JCDL '05 Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, 100–109.

Øverlier L./ Syverson, P. (2006). Locating Hidden Servers, In: 2006 IEEE Symposium on Security and Privacy, 100–114.

Papsdorf, C. (2013). Internet und Gesellschaft. Wie das Netz unsere Kommunikation verändert. Frankfurt a.M./ New York: Campus.

Pariser, E. (2011). The Filter Bubble: What the Internet Is Hiding from You. New York: Penguin Press.

Power, M. (2014). Life after Silk Road: how the darknet drugs market is booming. Retrieved from http://www.theguardian.com/technology/2014/may/30/life-after-silk-road-how-the-darknet-drugs-market-is-booming

Raghavan, S./ Garcia-Molina, H. (2000). Crawling the Hidden Web. Stanford: Digital Libraries Technical Report.

Rogers, E. M. (1962). Diffusion of Innovations. Macmillan Company: Free Press of Glencoe.

Rosander, M./ Eriksson, O. (2012). Conformity on the Internet – The role of task difficulty and gender differences. In: Computers in Human Behavior 28, 1587–1595.

Sabella, R. A., Patchin, J. W./ Hinduja, S. (2013). Cyberbullying myths and realities. Computers in Human Behavior 29, 2703–2711.

Selinger, E./ Hartzog, W. (2014). Obscurity and Privacy. In: Pitt, J./ Shew, A. (eds), Routledge Companion to Philosophy of Technology. Retrieved from http://ssrn.com/abstract=2439866

Tor Metrics (2015a). Direct users by country. Retrieved from https://metrics.torproject.org/userstats-relay-country.html

Tor Metrics (2015b). Top-10 countries by directly connecting users. Retrieved from https://metrics.torproject.org/userstats-relay-table.html

Tor Metrics (2015c). Unique .onion addresses. Retrieved from https://metrics.torproject.org/hidserv-dir-onions-seen.html

Tor Project (2015). What is Tor? Retrieved from https://www.torproject.org

Zheng, Q., Wu, Z., Cheng, X., Jiang, L./ Liu, J. (2013). Learning to crawl deep web. In: Information Systems 38, 801–819.

Zillien, N.(2009). Digitale Ungleichheit. Neue Technologien und alte Ungleichheiten in der Informations- und Wissensgesellschaft. Wiesbaden: VS.